Attorney Docket No. 18961.00

IN THE APPLICATION

OF

YUNQI TIAN

FOR A

TWO-LEVEL INTERNET SEARCH SERVICE SYSTEM

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

Attorney Docket No. 18961.00

TWO-LEVEL INTERNET SEARCH SERVICE SYSTEM

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/203,943, filed May 12, 2000.

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

The present invention relates to resource discovery on the Internet, especially to the functions currently performed by search engines and directories.

2. DESCRIPTION OF RELATED ART

Search engines play a vital role in finding things in cyberspace. The first of these search engines, Archie, maintained a database of approximately 1,500 host computers, which housed files accessible through the file transfer protocol (FTP) space of the Internet. FTP sites were feasible in the early days of the Internet when the number of users and host computers was relatively small. The next advancement in search engine technology produced

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

a more user friendly system called Gopher, a subject-based, menu-driven guide for finding information on the Internet. Gopher searched all the files located on a particular host; however, it was impossible to know if all information relevant to a search resided on one host. Visiting all 6,000 Gopher servers would be an incredibly time-consuming task. Therefore, the search tool Veronica was developed to search Gopher space.

With the advent of the World Wide Web, it became necessary to create a new tool for cataloging information found on this portion of the Internet. Unlike a real-world landscape, the contours of the Web constantly shift as sites come and go, necessitating constant 'reconnaissance' in order to provide users with an accurate map of its offerings. This "map", which is really a catalog, is at the core of the search engine. When a user inputs a query into a dialog box, the user is not searching the Web per se. Rather, the user is searching an index of the Web created and continually updated by the search engine.

The first Web search engine was World Wide Web Wanderer. It used an automatic search agent called 'robots' to track the Web's growth. Then there was a search engine called ALIWEB. ALIWEB didn't use a search agent. Instead, it asked people to write descriptions of their Web service and register at ALIWEB. ALIWEB then periodically retrieved information from registered Web servers and combined them into a search database. Soon robots-based search engines and editor-based directories became the popular tools for Internet search. The search agent works like a chain reaction. It

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

starts from one Web page and follows the out-links to other Web pages and then repeats the same pattern on each Web page it finds.

Early search engines focused on 'breadth first' and 'depth first' search agents. Referred to colloquially as 'robots' or 'spiders', these agents were set loose onto the Web to index HyperText Mark-up Language (HTML) files residing on the myriad of servers connected to the Internet.

The 'breadth' first approach works by 'hydroplaning' over the expanse of the Web, taking note of any hyperlink references found in a file, but deferring any deeper inquiry in favor of moving on to cover as much territory as possible. In contrast, a 'depth first' approach works by honing in on a site, dropping anchor and thoroughly exploring every pointer leading from the file, drilling down until the search agent finds a file with no links outward bound.

Search engines operate in a three-step procedure. First, specialized types of software, referred to as 'robots', 'spiders' or 'crawlers', go out and retrieve information about Web sites. The search engine either has found these Web sites itself or the Web sites (through their Webmasters, those in charge of Web site management) have asked to be indexed. Some search engines tend to "index" (record by word) all of the terms on a given web site. Some may index only the terms within the first few sentences, the web site title, or the site's metatag, which is not viewable on the actual page and contains a short summary description provided by

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

the site designer. Search engines must re-sample the web sites periodically to detect any change since last indexing.

The second step is to store the indexes in a database with ranking for each web page. The rank reflects the relevance of the web pages to certain keywords. A proprietary algorithm is used to evaluate the index. Since different search engines employ different algorithms, the ranking results are not consistent from all the search engines. When an Internet user types a keyword or keywords as search query, search engines retrieve indexed information which matches the query from the database. This last step completes the service of search engines.

Internet directories operate on a different principle. They require human editors to view an individual web site and determine its placement into a subject classification scheme or taxonomy. Once done, certain keywords associated with those sites can be used for searching the directory's database to find web sites of interest, or people can follow the structure of the directory to find the information located under the directory structure.

Coverage on the search engines and directories affects the Internet usage significantly. People are heavily relying on the search engines and directories to use the Internet. According to a recent study, two-thirds to three-quarters of all users cite finding information as one of their primary uses of the Internet and more than 98% of active Web users rely on the Internet to find reference material, 30% on a daily basis and a further 40% on a weekly basis. The major Internet search engines - HotBot, Northern

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

Light and AltaVista - individually catalog at most 16% of the Internet's sites. As the amount of web pages increase, the coverage in the past two years has showed a decline. Combined, the results from all search engines the total Internet coverage is only about 42%. Due to the cost and time in individually assigning sites to categories and the editorial policy used by directory companies, lack of coverage is also a problem for Internet directories.

Although some search engines companies (Google and Inktomi) claimed their coverages are over 1 billion Web pages now, there is more content than current search engine companies can cover. There are more than one million new pages everyday. There are more non-HTML-text contents, e.g. Adobe's portable document format (PDF) and formatted files and multimedia files created. Also there are many non-crawlable contents, such as sites that have no links pointing to them, sites screened by a login, corporate intranets, sites that use robots.txt scripts to bar search robots, and deep content. Studies show that the "invisible Web" contains deep contents, nearly 550 billion Web pages, and most of which are open to the public but never touched by search engines. It is estimated that more than 100,000 deep Web sites exist. Another reason that search engines have difficulty finding all information on the Web is the structure of the Web, a bow tie shape according to a recent study. There is a large cluster of the Web that contains Web pages that cannot be reached by links.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

Some researchers see the coverage problem as damage to the intention of the Internet as a public good. The Internet as a public community embodies the ideals of a liberal democratic society. It is a rich array of commercial, political, academic, and artistic activities that fosters associations and communications of all people around the world, and provides a virtually endless supply of information. As technology progresses, it is certain that there will be more Internet applications. If trends on Internet directories and search engines lead to a narrowing of options, the Internet as the kind of public good that many people envisioned will be seriously undermined.

The information retrieved from search engines doesn't satisfy relevance very well. The indexing methods used by current search engines often misrepresent the contents in the indexed Web sites. Web site builders don't have much control on what they want Web users to know about their Web site. To increase a Web site's chance to be indexed correctly and to be placed on a higher spot in the "found sites" list, Web designers need to spend extra efforts to make a Web site suited for the search engines. This is always a confusing job because each search engine uses a different algorithm for indexing, and many keep it a secret.

Since the relevance is poor, as Web users conduct searches by using search engines, they suffer so-called "information overload", i.e. too much irrelevant information and no efficiency. In the worst cases, submitting broad query terms to search engines can result in literally hundreds of thousands of potential Web pages

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

identified. Many times users also get the same Web site and/or pages repeatedly appearing in the found result. To find what they want, users usually need to visit several search service Web sites.

Recency is poor. 50% of Internet users cite as one of their typical search problems as searches that turn up broken links. The bigger the search engine service is, the higher percentage of the dead links it has. It seems that there is a trade off between comprehensiveness and recency. Reducing the time between re-sampling is a big challenge for search engines. It will also unreasonably increase a visited Web site server's load. There is a considerable backlog on the directory service; for example, it can take six months for Yahoo! to put a site under its directory, if the editor decides the Web is suitable to be included. Therefore, recency will be a serious problem as the Web increases with a fast speed.

For current search engines and directories, Web users don't have much to say at getting a better search service. Because the Internet grows so rapidly, a self-improving search service is necessary for Web users.

Metadata, structured data about data, as a way to improve Internet searching has been proposed by Dublin Core Metadata Initiative since 1995. World Wide Web Consortium (W3C), under the leadership of Tim Berners-Lee, also proposed Resource Description Framework (RDF) for broader Internet applications including resource discovery. However, these metadata standards have to be recognized by search engines. Currently without support by any of

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

the major search engines, there is no reason for Web site builders to put them into the Web pages and the Internet community cannot benefit from them.

The related art includes articles that call for improvements in searching on the Internet. Searching the Web: General and Scientific Information Access and Accessibility of Information on the Web by Steve Lawrence et al., and Defining the Web: The Politics of Search Engines by Lucas Introna et al. discuss limitations of current search engines.

Inventions of interest, as depicted in patents, include U.S. Patent Number 5,283,731, issued on February 1, 1994 to James E. Lalonde et al., which describes a computer based classified ad system and method.

U.S. Patent Number 5,319,542, issued on June 7, 1994 to John E. King, Jr. et al., describes an electronic catalog ordering process and system.

U.S. Patent Number 5,649,186, issued on July 15, 1997 to Gregory J. Ferguson, describes a system and computer based method providing a dynamic information clipping service.

U.S. Patent Number 5,721,910, issued on February 24, 1998 to Sandra S. Unger et al., describes a database system and a method of producing a database which can be used to assign scientific or technical documents, such as patents and/or technical or scientific publications and/or abstracts of these patents or publications, to one or more scientific or technical categories within a multidimensional hierarchical system which reflects the business,

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

scientific or technical interests of a business, scientific or technical entity or specialty.

U.S. Patent Number 5,727,156, issued on March 10, 1998 to Dirk Herr-Hoyman et al., describes a method and apparatus for posting hypertext documents to a hypertext server so as to make the hypertext documents accessible to users of the hypertext document system while securing against unauthorized modification of the posted hypertext documents.

U.S. Patent Number 5,745,882, issued on April 28, 1998 to Matthew J. Bixler et al., describes an interface for an electronic classified advertising system that includes the capability for the user to enter search criteria for an item of interest, to save the search criteria and to be notified by the system when an item matching the search criteria is entered into the system.

U.S. Patent Number 5,794,236, issued on August 11, 1998 to Joseph P. Mehrle, describes a computer based system that will classify a legal document into a location within a legal hierarchy.

U.S. Patent Number 5,799,284, issued on August 25, 1998 to Roy E. Bourquin, describes a computer system that utilizes client/server software to allow users of the client software to log into a server and publish information about a product or service.

U.S. Patent Number 5,855,013, issued on December 29, 1998 to Dave C. Fisk, describes a method and apparatus for creating and maintaining a computer database using a virtual index system.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

U.S. Patent Number 5,870,717, issued on February 9, 1999 to Charles F. Wiecha, describes a system for ordering items over a computer network using an electronic catalog.

U.S. Patent Number 5,963,951, issued on October 5, 1999 to Gregg Collins, describes a computerized on-line dating service that provides user-controlled perusal of search results.

U.S. Patent Number 5,974,409, issued on October 26, 1999 to Sankrant Sanu et al., describes an enhanced find system and method for locating offerings within an interactive on-line network.

U.S. Patent Number 6,009,410, issued on December 28, 1999 to Suzanne L. LeMole et al., describes a method and system for presenting customized advertising to a user on the World Wide Web.

International Patent document WO 98/19417, published on May 7, 1998, describes an integrated computer-implemented corporate information delivery system.

None of the above inventions and patents, taken either singly or in combination, is seen to describe the instant invention as claimed.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

SUMMARY OF THE INVENTION

The present invention is a two-level Internet search service system. An Internet search service system should have five basic features to meet Web users' needs: comprehensiveness, recency, relevancy, efficiency, and self-improvement. The two-level Internet search service system according to the invention is targeted to meet all of them and will provide a more superior search result relative to current search engines and Internet directories. The two-level Internet search service system can provide a feasible way to keep up with the growing speed of the Internet. The two-level Internet search service system can also reduce a Web site server's load by eliminating search engines' visits. Within the platform of the two-level Internet search service system, Web site builders or Webmasters will have better control on correctly presenting indexes of their contents; Web users will have better control to clarify what they are looking for, can give feedback on service improvement, and can give their opinion for which Web site is more relevant to their query. By using this two-level Internet search service system, eventually it will be possible to provide a one-stop search at one Web site instead of going to multiple Web sites in a current situation.

In the two-level Internet search service system the searches are distinguished into two levels: search service provider (SSP) level and in-site level. The objective of SSP level search is to bring Web users to the right Web destination. The objective of in-

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

site level is to find exact information Web users are seeking. At SSP level three tasks need to be accomplished. The first is getting the correct search indexes about Web sites, which is the base for other tasks; the second is organizing the indexes at the search service's Web site; and the third is helping people search efficiently. The three components of SSP level search service system include (1) input system, (2) data organizer at the search service's Web site, and (3) interactive search service provided by the SSP.

At in-site level three tasks need to be accomplished. The first is improving Web units' navigation system; the second is implementing metadata standards; and the third is installing an in-site search engine to utilize the metadata. Three components are needed for in-site level search service: (1) a design aid system, (2) an authoring tool, and (3) an in-site search engine tool kit.

Accordingly, it is a principal object of the invention to provide a two-level Internet search service system comprising Web unit submission means, detecting and sorting means, multi-directory and database means, interactive search service means, and in-site search means.

It is another object of the invention to provide a software means for automatically generating standard index data and multi-directory entries and sending them back to SSP's server.

It is another object of the invention to provide a data organizer at SSP's server to process and store the data from submission.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

It is another object of the invention to provide software means for interactive search activities among Web users, Web unit builders or Webmasters, and SSP in both SSP level search and in-site level search.

It is an object of the invention to provide improved elements and arrangements thereof in a two-level Internet search service system for the purposes described which is inexpensive, dependable and fully effective in accomplishing its intended purposes.

These and other objects of the present invention will become readily apparent upon further review of the following specification and drawings.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of the SSP level search service system according to the invention.

Fig. 2 illustrates the internal structure of a web site.

Fig. 3 is the flow chart of the sorting and detecting program according the invention.

Fig. 4 describes how a web unit can be placed in different categories of directories and indexed.

Fig. 5 is the keyword or phrase search procedure flow chart according to the invention.

Fig. 6 is a block diagram of second function of search service at SSP level.

Fig. 7 is a block diagram of in-site level search service according to the invention.

Similar reference characters denote corresponding features consistently throughout the attached drawings.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In the description of the preferred embodiment, reference is made to the accompanying drawings which form a part hereof, and in which is shown by way of illustration the specific embodiment in which the invention may be practiced. It is to be understood that other embodiments may be utilized as structural changes may be made without departing from the scope of the present invention.

The present invention is an Internet search service system. When an Internet user retrieves Web pages, they use a browser to transmit HyperText Transfer Protocol (HTTP) commands from their computer to a Web server executed by a connected computer. In turn, the Web server responds with an HTML (or other formatted) page that is transmitted to the browser for display to the user.

Typically, users access Web pages by using a SSP (Yahoo.com, AltaVista.com, etc.) to find pages regarding a topic of interest. If the Web page is of some interest to the users, they "bookmark" the HTTP Uniform Resource Locator (URL) for that page in their browser in order to easily find the Web page in the future.

Fig. 1 is a block diagram of an exemplary hardware environment of the preferred embodiment of the present invention, and more particularly, illustrates a typical distributed computer system, wherein client computers, or users, are connected via a network to server computers, or sites. A typical combination of resources may include clients that are personal computers or workstations, and servers that are personal computers, workstations, minicomputers,

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

and/or mainframes. The network preferably comprises the Internet, although it could also comprise intranets, extranets, local area networks, wide area networks, etc. As shown, Web sites **10** are accessed by Web users **12** via the Internet. The Web users **12** may utilize the services provided by the inventive Internet search service system **20** at SSP level. The Internet search service system **20** includes a sorting and detecting means **22**, search service means **24**, which includes a server **28**, and directories and database means **26**.

Each of the computers, be they client or server, generally include a processor, random access memory, data storage devices, data communications devices, a monitor, user input devices, etc. Those skilled in the art will recognize that any combination of the above components, or any number of different components, peripherals, and other devices, may be used with the client and server.

For the purpose of indexing, the Internet search service system **20** is based on a new concept of the Internet. Currently, people use Web sites, Web pages, and Web documents to describe the information entities on the World Wide Web. Basically directories use Web sites as the search result, while search engines use Web pages. The proper way to index and retrieve information is to use a Web unit as the information entity.

A Web site is the collection of total contents of a Web entity and has only first level URL, e.g.

http://www.fastcompany.com/

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

http://www.mlb.ilstu.edu/

http://movies.yahoo.com/

For some Web sites, which are hosted by a hosting Web site, the URLs look like

http://maxpages.com/dbnursery

A Web unit is a self-contained and distinguishable sub entity of the Web site. Its URL has several levels. Usually a navigation structure is built for a Web unit so that its users can go around the Web unit in a manageable and easy manner. Examples of Web units are:

http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html

http://hometown.aol.com/algen4me/page/index.htm

http://www.lakewoodconferences.com/wp

http://www.ncsa.uiuc.edu/demoweb/

http://www.neci.nj.nec.com/~lawrence/science98.html

Here a personal home page, the second example, is treated as a Web unit, because it is not big but needs to be indexed. A research paper, which may be a single Web page as the last example shows, is also treated as a Web unit. A Web site may have several Web units in it or can be treated also as a single Web unit if the Web site is not very large.

A Web page is an element of the Web unit. It has the maximum levels of URL within the Web site or Web unit. A Web unit (even a Web site) may only have one Web page, but generally a Web unit has several Web pages. Examples of Web pages are

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

http://www.lakewoodconferences.com/wp/introduction.html

http://www.lib.berkeley.edu/TeachingLib/

The second example is called an intermediate Web page. An intermediate page is a bridge to more Web units.

Fig. 2 shows a typical internal structure of a Web site **30**. The Web site **30** has a home page **32**. From the home page **32** the navigation structure brings people to Web units **36** or an intermediate Web page **34**. From the intermediate Web page **34** people can reach more Web units **36** at a lower level. Web units **36** can have one page or several pages. Within a Web unit **36** there may be more complicated structures. There are also some hyperlinks among Web units **36** or between two Web pages in different Web units **36**. **Fig. 2** doesn't show the hyperlinks, because they are not critical for the indexing purpose.

The invention uses a Web unit as the core element to catch the identity and functionality of a Web site. Only indexing the entire Web site will risk loosing too much information. Many Web sites can contain over hundreds or even thousands of Web pages. A brief description in the directory won't work well to tell what can be found in a Web site. Indexing each page is not necessary because within a Web unit all the pages are related and integrated into a whole. Therefore, the two-level Internet search service system indexes Web units at SSP level and directs Web users to matching Web units. Once Web users enter a Web unit they should be guided by the in-site searching means, which is described later, and easily find what they are looking for.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

The process of SSP-level search is like following: Web unit builders or Webmasters come to the SSP's Web site and register at the input Web unit. Then they download submission software from the SSP **28**. The software contains a tutorial for submission, an index form, a multi-directory entry form, a suggestion form, and an update reminding program. They fill in the index form, multi-directory selection form, and suggestion form and send them back to the SSP's server. At the SSP's server, the input information is checked via SSP's editing standard, sorted, and stored. When Web users come to the SSP's Web site for an Internet search, they can follow directory structure or send a word query to SSP's search engine. SSP uses submitted information to help Web users narrow the search scope. Also SSP gives Web users the opportunities to fill feedback, post unsolved problems, rate the Web units they have visited, and report any error they find in their search process. While Web users are searching the Web units, SSP also provides personal directory service and personal agent to improve the efficiency of search at in-site level.

As shown in **Fig. 3**, the first component at SSP level search of the Internet search service system is self-submission. The key is a new input form, which is used to catch all the vital information about a Web unit. There can be as many as twenty information items needed to catch the identity of a Web unit. Conventional search engine and directory practice only ask for one URL and one e-mail address from each Web site. Some also ask for a brief description of the Web site. But current search engines and directories will

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

never obtain the information gathered by this method. The following table shows some examples of information items needed for submission. Not all the items are necessary for every Web unit and also the list may be expanded to improve the accuracy.

Submission Items Table

1. Title
2. URL of the Web unit
3. URL of the parent Web site
4. Brief description
5. Primary keywords and phrases
6. Secondary keywords and phrases
7. Total number of the Web pages in the Web unit
8. Entity name (company, organization, or person) and geographic location
9. Off-line companion for not-pure-Internet entity
10. Author's name (for academic papers and knowledge materials)
11. Product or category names and brands
12. Targeted users (geographic location and user group - age, gender, occupations, and etc.)
13. Membership information
14. Specifications - size, text based, multimedia based, and etc.
15. Language information - Chinese, Dutch, English, French, German and others

16. Special requirements needed to view the Web unit

17. Geographic location of Web site

18. Update information

19. Contact information

20. Other special features

Web unit builders or Webmasters will download a submission software means **38** from search service **24**, which contains a guide and registration form for submission **40** process. The submission program **38** contains a tutorial, a Web unit index form, a multi-directory entry form, a suggestion form, and an update reminding software. The tutorial will explain the web unit concept, index form, directory structure, and relationship between indices and directory entries. The index form contains blanks for Web unit builders or Webmaster at Web site **10** to fill in index items. The multi-directory entry form contains the directory selection structure for Web unit builders or Webmasters at Web site **10** to select categories, branches and nodes. They can select as many locations in the directory structure as they think it is suitable, as shown in **Fig. 4**. Both the index form and the multi-directory entry form will automatically generate standard data format to be sent to SSP's server. The suggestion form is used for Web unit builders or Webmasters at Web site **10** to suggest new branches and nodes in the directory structure, because most likely the initial design of the directory structure may need redefining and reconstructing to reflect the rich and complex Web contents. The

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

updating reminding software will stay at Web site **10** and automatically detect new changes in the site's contents. Whenever there is a change, it will generate a reminding message to Web unit builders or the Webmaster to submit the new information. The basic features of comprehensiveness and recency can be achieved through this input system. Also the database at SSP's server will be more condensed and richer on information for covering the same amount of Web sites than conventional search engines' databases.

The second component at SSP-level search of the Internet search service system is data organizer. The information or Web unit input **62** gathered from Web unit builders or Webmaster at Web site **10** will go through sorting and detecting means **22** and stored in database **26**. The database **26** can be further divided (by sorting and detecting **60**) into a categorized directory **66** (seen in **Fig. 4**) to organize the Web units and an index database **64** to store all the input information. Human editors are also involved in the data organizing activities.

The sorting and detecting program works as shown in **Fig. 3**. After the Internet search service system **20** receives a submission, the sorting and detecting program will go through several checking steps (**42**, **44**, **46**, and **48**) to find whether there is an error in the submission form **40**, i.e. errors in the index form, misplacement in directory entry, or other errors that violate the SSP's editing policy. Any error will be recorded in **58**. After all the checking is finished, **50** will check whether there is any error recorded at **58**. If there is an error, a return form **56** will be sent to the

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

submitting Web site to ask for a re-submit. Otherwise, a sorting program **52** will sort and send the data to a database **54**. One thing which can be controlled by the detecting is the ratio **44** of primary and secondary keywords and phrases. The rule-of-thumb for the ratio is less than 1/3, i.e. every one primary keyword should have at least three secondary keywords. For any submission which has more than fifteen keywords or phrases, the keywords or phrases must be divided into primary and secondary. If the ratio is not reasonable, a resubmission is required. A limit is placed for the length of the description **46**. Another possible limit is the number of product names **48**. If number of product names is over a certain amount, the category name is required to replace the product names.

There may be more checking items as the input form becomes more complex. Also the ratio of primary and secondary keywords and the limit on only submitting primary keywords will be adjustable for accuracy. Web users will play a role in detecting errors in submitted information, as described later. Their contributions will be integrated into the detecting process.

The directory structure is categorized into information, online shopping, online service, entertainment, business-to-business, and community, as shown in **Fig. 4**. As the Internet grows, there may be new categories added to the Internet search service system **20**. Under each category, a complete directory forms a tree structure with branches and nodes. Therefore, there are as many directories as categories. A Web unit

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

may be placed in several directories. There will be some submissions that cannot clearly identify where the Web units belong. In this case the Web unit builder or Webmaster can suggest a new directory branch and node. The editors at the Internet search service system will use their expertise to place those Web units and modify the directories. In the index database **64**, the same keyword or phrase may have different ranks, depending on whether the Web unit builders or Webmaster submitted them as primary or secondary.

The third component of the Internet search service system **20** is the search service **24**. The primary search methods are following directory structure and using keyword or phrase query. As the input system gets more and more information from Web unit builders or Webmasters, the search can also have other forms. The search process is an interaction between SSP and Web users. The search service **24** uses information collected to help Web users narrow down the result to a condensed and relevant Web units list. It provides a personalized directory system and a personal agent to Web users. It also provides a platform for self-improvement.

The first function of the search service is the keyword or phrase search, which is shown in **Fig. 5**. When the information retrieved **70** exceeds a certain amount, the search service program **24** will pop out question **72** to narrow down the search scope and help make the found result more relevant to user's query. Here, most of the information gathered from the self-submission **40** is used as the variable items **76** to eliminate non-related Web units

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

from the found list **70** and generate a new narrowed list **78**. For example, for one search query there are 200 matching Web units, but the "target users" from all 200 Web units are different.

This information can be used by a user to decide which kind of "target user" he/she is. Then all the Web units that match his/her choice of "target user" will form a new and narrowed list. This process is continued until the user feels comfortable to look at the list. Then they will enter **74**, where they can also get help from a search agent described later. The features of relevancy and efficiency of the invention are clearly shown. The found results will follow a rank to be shown to the user. The matching between search query and primary keyword/phrase will have a higher rank than secondary keyword/phrase does. User ratings, described below, will also be used to improve the ranking. Through the first function of the search service, the relevancy issue of Internet search can be improved dramatically.

As shown in **Fig. 6**, the second function of the search service is to provide a personal directory **84** and search agent **88** to Web users **12**. The personal directory **84** will use multi-directory structure to record Web units and/or Web pages that have been visited by individual Web users. The personal directory can help the Internet search service system **20** set up a profile for individual Web users. The Internet search service system **20** will use this information to recommend Web units to Web users. It will also use updated information from the self-submission process to automatically update (by update and analysis software **90**) the

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

personal directories for Web users and inform them of the changes instantly. The update and analysis software **90** also analyzes search patterns of individuals and gives suggestions for search improvement and recommends Web units, as noted above.

Web users can use the search agent **88** in two different ways. They can start from the found list **86** after the narrow down process or use personal directory without a new search. Personal agent will help Web users at the in-site level search. It can initiate in-site search engines at Web sites **10** with a Web user's query. It can automatically go to other Web units in the visiting list, which forms from either found list **86** or in the same branch of the personal directory **84**, to find related Web pages while the user is in the current Web unit. The search agent together with the first function will greatly increase the search efficiency.

This relationship is indicated by the arrow between **84** and **88**. This arrow indicates means to allow communication between the search agent and the personal directory so that the search agent can carry search requests to the Web units in the personal directory to initiate the in-site search engine tool kit for a more specific search. Optionally, the search agent can act as a Web robot to search a Web unit following the navigation structure built into the Web unit

Another important function of the user's service is self-improvement. Web users **12** give feedback, rating scores, and requests to the system **20**. Several aspects are included in the self-improvement process:

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

Relevancy rating and detecting - Web users can rate the relevancy of the Web unit that they are visiting. They can also report any error or spam in the data collected about the Web unit. Their rating scores will be integrated into the index database **64** to improve the ranking accuracy and any error or spam detected will be corrected through the editorial procedure of the system;

Supporting community - users with similar search interests can use this channel to exchange experiences and help each other on the search;

User feedback - Web users can make suggestions for search service improvement. When the feedback is transmitted to the Web unit builders of Webmasters, it may include improvement suggestions from SSP experts; and,

Unsolved query posting - any unsuccessful search will be recorded and posted to ask for help from the Internet community. A Web unit will be utilized for posting unsolved search problems related to Internet search and Internet use and for providing solutions to the posted problems. Any unknown word will also be posted to ask for explanation and references. By doing that, the inventive Internet search service system **20** will gradually cover every word and phrase that people are searching for. This is another way to improve the comprehensiveness of the Internet search. By involving content providers and seekers, the Internet search service system can also solve the critical problem faced by current search engines and Internet directories, i.e. keeping up with the Internet's growing speed.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

At the in-site level search, Web users can follow a navigation system built inside of the Web unit or use an in-site search engine. The personal search agent can help them enhance search efficiency by searching other Web units on their visiting list while they are surfing in the current Web unit.

As shown in **Fig. 7**, at the in-site level search service the Internet search service system **20** performs three tasks to improve the search at Web sites **10** for Web users **12**. First, it uses Web users' feedback **104** to help Web unit builders or Webmasters at Web sites **10** to design a better navigation system by using design aid **106**. Second, it implements metadata standards by letting Web unit builders or Webmasters use an authoring tool **108**. Third, it installs an in-site search engine **110** to enhance the search capability for those Web sites or Web units that have used metadata.

The first component of in-site search service is the design aid **106**. Web users' feedback **104** is the major source for reconstruction of the Web site or unit, while other sources may also provide hints for improvement. The feedback will be in two forms. The first one is the generic feedback form that can be used for any Web unit or Web site. The other one is the customized feedback form that meets special needs for certain Web sites or Web units. By collecting and analyzing the data, the Internet search service system **20** will generate a suggestion form for Web sites or Web units. Then Web unit builders or Webmasters will work with experts in the Internet search service system **20** to reconstruct the

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

Web site or Web unit. Web unit builders or Webmasters can also choose using designing tools and testing tools provided by the Internet search service system **20**. The improved site or unit structure and navigation system can help Web users to surf around and find information more easily when they next visit the Web site or Web unit.

The second component of in-site search service is the authoring tool **108**. Web unit builders or Webmasters will download this tool. The tool employs software means to make the metadata writing easy. It will use simple forms to let users fill in the data and then generate standard data format for the in-site search engine **110** to use.

The third component of in-site search service is the in-site search engine tool kit **110**. This tool kit will integrate in-site robots, databases, and search engines for Web unit builders or Webmasters. The tool kit can be a generic one or a customized one, depending on the needs of Web unit builders and Webmasters. The in-site search engine has the capability to communicate with Web users' personal search agent. When Web users use the personal search agent to help them on the search, the agent will carry their query or commend and travel to the Web site or Web unit on the visiting list. When it gets there it will find the in-site search engine and convey the search request. The in-site search engine then will start the search and bring the result back to the Web user's browser. Web users can also initiate in-site search engine themselves.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000

The in-site level search is very focused and efficient. It builds the search mechanism on Web users' needs and the unique purposes of the individual Web site or unit. The resources description and retrieve are localized. Navigating the Web unit is easy. The in-site search engine works fast on a relatively small database. Web users will have more satisfied search experiences from an improved and user-friendly in-site search.

It is to be understood that the present invention is not limited to the sole embodiment described above, but encompasses any and all embodiments within the scope of the following claims.

LITMAN LAW
OFFICES, LTD.
P.O. BOX 15035
ARLINGTON, VA 22215
(703) 486-1000